The Baxter Return of Experience on the Use of Association Rules to Construct its Product Line Model

Alberto Lora-Michiels, Camille Salinesi, Raul Mazo

The Baxter Return of Experience on the Use of Association Rules to Construct its Product Line Model

Alberto Lora-Michiels[1, 2] Camille Salinesi[2] and Raúl Mazo[2,3]
[1] *Baxter International Inc, Belgique*
[2] *CRI, Panthéon Sorbonne University, 90, rue de Tolbiac, 75013 Paris, France.*
[3] *Ingeniería de Sistemas, Universidad de Antioquia, Medellín – Colombia.*
albertoloram@gmail.com, camille.salinesi@univ-paris1.fr, raulmazo@gmail.com

Abstract

A very promising approach to increase productivity, quality and competitiveness approach of information systems development is the reuse and development of a family of systems guided by Product Line (PL) practices. One of the main goals of PL engineering is to develop a model that represents the family of products (product line model PLM), which is then customized to configure individual products.

The successful definition of PLMs that accurately represents the information in the requirement specifications still depends heavily on the intuition and experience of the software architect. Our work provides assistance for this process. We have developed a semi-automated method to construct product line models based on collection of related artifacts or existent products models as a result of a feature mining process.

The approach is evaluated using bill of material as a collection of product models to develop and construct a constraint based PLM. The performance of our method is calculated by estimating the time complexity and constructing the PLM for different random samples of 536 products in Baxter Bioscience. More than 92% of the relationships are properly predicted only by using 75% of the total available products.

1. Introduction

Approaches to construct PLMs are often focused on using clustering methods to elicitate, prioritize and triage requirements. Rather than a systematic process, the derivation of an initial product line model from the provided requirement descriptions has remained something of a black art and almost on domain analyst experience. In our approach, once the product model for each application is defined we propose a semi-automated method to guide and build a PLM. In this process, we use a data mining association rules technique (Apriori algorithm [Agrawal *et al.*, 93]) and independence tests. Our approach is based on a sequence of logic activities to achieve the identification of some relationships in a collection of product models. We are interested in the identification of structural, transversal and group cardinality associations.

The rest of the paper is organized as follows: Section 2 briefly overviews our approach by identifying tits main aspects. Section 3 presents the results obtained through a real study case. Section 4 concludes the paper and describes future works.

2. Approach

To begin with our approach it is necessary to dispose of a collection of related features or artefacts for each application. Artefacts or features could be extracted from repositories and by means of clustering process the hierarchical relation could be established.

The main phases in PLM construction are to represent the structural dependency, transversal dependency, and grouped cardinality.

First of all, the structural dependency deals with a parent child representation and of course a set of bundles. Parent child relationships and bundles are in the most of the cases obtained thanks to a clustering [Chen *et al.,* 05]. The mainly objective is not only acquiring the right parent child relationship but also to figure out those optional and mandatory relationships.

Second, the transversal dependency studies the behaviour among features that are not member of the parent child link. The principal goal of this transversal dependency study is in exploiting:

(i) All the possible mutually exclusive relationships (perhaps for a different bundle member or between brothers in the same bundle). Those mutually exclusive transversal relationships are known as *excludes* relationships.

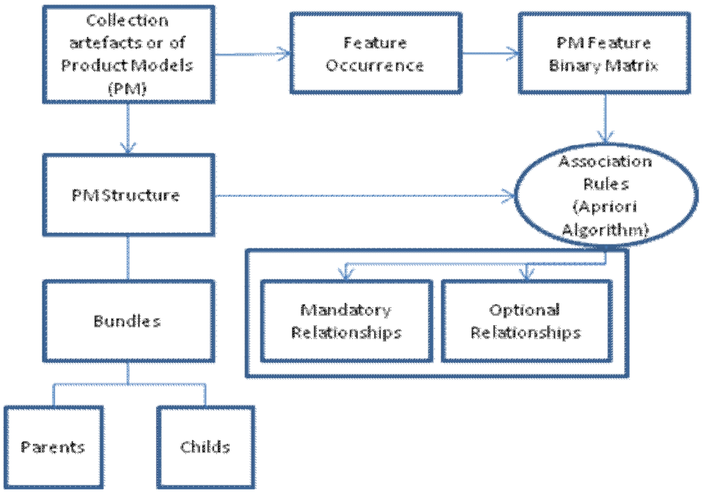
(ii) All the possible relationship dependencies between members of different bundles or brothers of the same bundle. For instance the selection of a specific feature may require the selection of another feature. Those types of relationships are called *requires.*

Third, the group cardinality study. The concept of groups was further generalized in [Riebisch *et al.,* 02] [Czarnecki, 98]: as a set of features annotated with a cardinality specifying an interval of how many features can be selected from that set. Thus, group cardinality is a property of the relationship between a parent and a set of optional sub features.

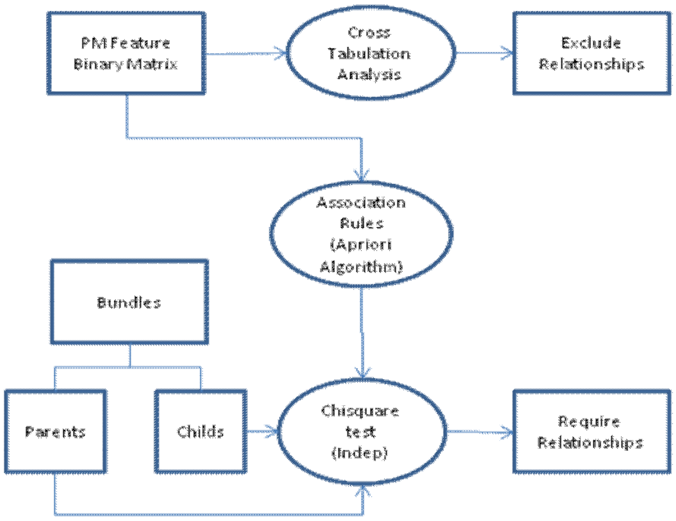
The three main phases of our approach are depicted in figure 1: structural, transversal and cardinality analysis. The core and original contributions of this work are represented with round circles in the figure:

- ✓ cross table analysis to determine exclude-type relationships;

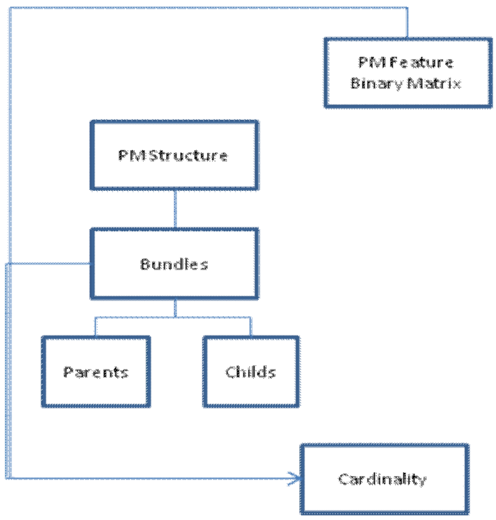
- ✓ association rules analysis used to identified mandatory and optional-type relationships;
- ✓ chi-square independence test combined with association rules to identify require-type relationships.



a) Structural analysis



b) Transversal analysis



c) Cardinality analysis

Figure 1. PLM construction framework [Lora-Michiels, 09]

The result of this method is to identify all the possible relationships while using constraint programming [Salinesi *et al.*, 09]. With this collection of constraints and by passing through a refinement process it is possible to obtain an accurate product line model

3. Baxter Bioscience study case

Baxter International Inc. develops, manufactures and markets products that save and sustain the lives of people with hemophilia, immune disorders, infectious diseases, kidney disease, trauma, and other chronic and acute medical conditions.

To construct the packaging product line model in our society, we focused our study around all the components that constitute the packaging process of the different treatments that Baxter Bioscience produces. We have worked with 536 packaging bill of materials (BOM) as product models and we have also handled more than 1500 items. After generalizing items, we proceed to apply our approach and evaluate the results obtained by estimating the algorithm time complexity and the scalability generating the desired constraints. First, examining the time complexity of the algorithm that supports our approach, we have observed that it is really efficient but it presents some limitations when studying group cardinalities. Group cardinality identification is a process that takes more time than other ones (bundles generation, mandatory, optional, requires and excludes relationships identification). Second, performing a paired comparison of constraints generated from different random products samples (Figure 2). We can observe structural dependencies show a high predictive capacity:

> 95% of the mandatory and optional relationships are founded when we take a random sample size of at least 350 products. The totality of the mandatory relationships are then discovered when the random sample size is greater than 450 products.

However excludes and, especially, requires relationships, seem to depend to the problem size that is it, the number of constraints increases when sample size increases. This can be explained by examining the nature of the data used in our study case. Structural relationships mainly depend on the composition of the product; thus they depend of the parent child relationships or BOM composition and transversal dependencies are related to relationships attributes. More products means more attributes, and at the end, this means that more transversal relationships to be discovered.

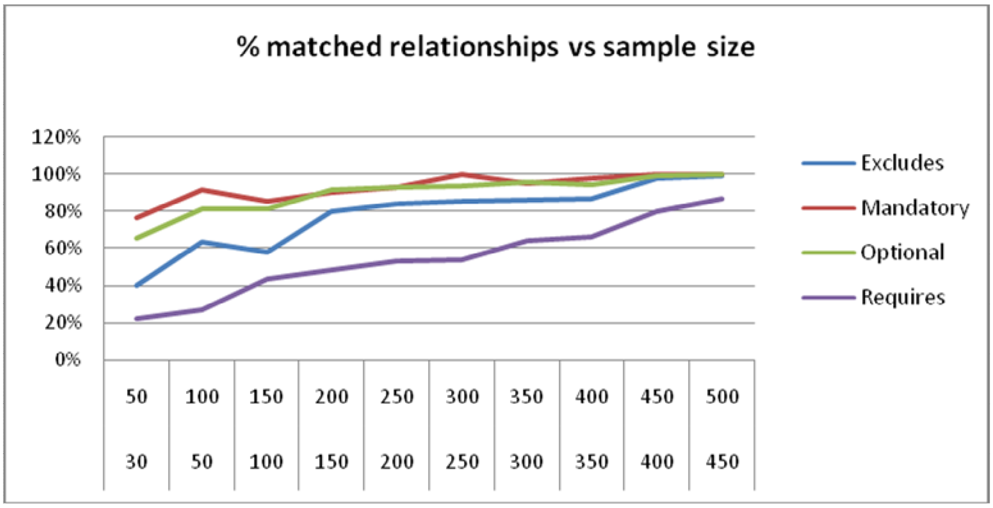


Figure 2. Relationships matching (different sample size comparison)

4. Conclusions

Our work is one of the first real scale experience of automation of the construction process of PLMs. To our knowledge, it is the first approach that integrates statistical techniques to identify commonalities and variabilities in a collection of a non predefined number of product models. Our approach can be used both for academic purposes as well as in industry. Indeed, although rigorous, our proposal needs to be expanded and benchmarked with respect to alternative strategies explored, and implemented into a marketable tool.

Our experience showed that there is a need for a method that is able to deal with richer input information. For example, we had products that are defined with more complex than Boolean-type features, as for instance scalar variable (*e.g.* integer or real values as in performance characteristics of systems) or set variables (when system features can be instantiated a varying number of times in the same products). As a consequence, we believe that more complex relationships can be needed in the target PL models. How can these be specified?

Several other fundamental questions came to our mind while we were designing and experimenting our method. For instance: what is a good quality model to construct a product line model? How to deal with ambiguous information to construct a product line model? How to deal with more complex constraints? What statistical tools could be used to support the aforementioned questions?

5. References

R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In Proceedings of the 1993 International Conference on Management of Data (SIGMOD 93), pages 207--216, May 1993.

K Chen, W Zhang, H Zhao, H Mei. An Approach to Constructing Feature Models Based on Requirements Clustering. Int. Conference on Requirements Engineering. Paris, 2005.

Czarnecki, K. Generative Programming PhD thesis, Technische Universitat Ilmenau, Germany, 1998.

Lora-Michiels, A. Product Line Model Construction based on association rules: A Statistical Approach, Master Tesis, Université Panthéon Sorbonne, Paris - France, September 2009.

Methods, Techniques and Tools for Product Line Model Verification. Centre de Recherche en Informatique's Research Report. Paris France, October 2008.

Salinesi, C, Rolland C., Diaz D., Mazo R. Looking for Product Line Feature Models Defects: Towards a Systematic Classification of Verification Criteria. In 17th IEEE International Requirements Engineering Conference, Atlanta USA, September 2009.

Riebisch M., Böllert K., Streitferdt D., Philippow I. Extending Feature Diagrams with UML Multiplicities, 6th Conference on Integrated Design and Process Technology, Pasadena, CA, 2002.

Biographie

Alberto Lora Michiels: Industrial Engineer 1999 (Escuela Colombiana de Ingenieria), Certificate in Statistics (Universidad Nacional de Colombia 2005), M2 en Système d'information et décision (Univeristé de paris 1). Currently, Material Requirement Planner Leader Supply Chain Baxter International Inc Lessines Belgique. Research interests: Product line model construction and data mining.

Camille Salinesi is a lecturer at Sorbonne University. He has obtained his PhD. degree in Computer Science at Paris 6 University (1999, France). His research interests are in the areas of requirements engineering, product line engineering, ERP systems, IS and business alignment and change management.

Raul Mazo Peña is a System Engineer at University of Antioquia (2005, Colombia). He has a MSc. at Panthèon Sorbonne University (2008, France) and at present is a PhD. candidate in Computer Science at the same university under advisor of Dolette Rolland and Camille Salinesi. His research interest are in the areas of product line model verification and validation, ERP systems and network security.